

Data Visualization of the Brazilian National High School Exam: VisDadosEnem

Visualização de Dados do Exame Nacional Brasileiro do Ensino Médio: VisDadosEnem

César H. C. Santos¹, Maykon C. Freitas¹, Robson R. Lemos¹, Alexandre L. Gonçalves¹
(1) UFSC, R. Pedro J. Pereira, 150. Araranguá, SC, Brasil, {robson.lemos, a.l.goncalves}@ufsc.br

Resumo:

Hoje em dia, baseado na quantidade de dados disponíveis surgem situações em que a visualização da informação torna-se fundamental para o entendimento e interpretação dos dados. Dentro deste contexto, este estudo tem como objetivo relacionar e explorar, através de técnicas de visualização, dados abertos educacionais em um contexto multidisciplinar. Para tal, foi desenvolvido uma aplicação Web para visualização de dados do exame nacional Brasileiro do ensino médio (ENEM). Neste estudo utilizou-se os microdados do ENEM, a partir do portal Brasileiro de dados abertos. Como resultado das análises comparativas realizadas sobre os dados, por estados, regiões e tipos de escolas, observou-se quais estados obtiveram as maiores e menores médias do ENEM em relação as médias nacionais. Por exemplo, no estado de Santa Catarina os alunos oriundos de escolas particulares obtiveram 12.70% acima da média nacional para este tipo de escola e a menor média de idade dos alunos inscritos correspondendo a 21,6 anos de idade. A aplicação Web de visualização possibilitou uma análise comparativa sobre o grande conjunto de dados e a exploração de possibilidades de visualização que seriam dificilmente possíveis de realizar por meio de representações tradicionais.

Palavras-chave: Visualização da Informação, Ciência dos Dados, Dados Abertos, Educação

Abstract:

Nowadays, based on the amount of available data, there are situations in which the information visualization becomes fundamental for the understanding and interpretation of the data. In this context, the study aims to relate and explore, through visualization techniques, open educational data in a multidisciplinary context. For that, a Web application was developed for data visualization of the Brazilian national high school exam (ENEM). In this study, the ENEM microdata was used, from the Brazilian open data portal. As a result of the comparative analyzes performed on the data, by states, regions and types of schools, it was observed which states obtained the highest and lowest averages of ENEM in relation to national averages. For example, in the state of Santa Catarina, students from private schools obtained 12.70% above the national average for that type of school and the lowest average age of enrolled students corresponding to 21.6 years old. The visualization Web application made possible a comparative analysis on the large dataset and the exploration of visualization possibilities that would be difficult to perform through traditional representations.

Keywords: Information Visualization, Data Science, Open Data, Education

1 Introduction

Based on the technological advances in data storage and retrieval there is a massive increase in the information available to any type of user through a web browser. There is a large volume of information available on the Internet, but the data are irrelevant when there is no meaning. In order to get the information, the data must be interpreted and related so that they are presented within a context and allow the generation of knowledge. In this way, developers have been paying special attention to data visualization. According to Freitas et al. (1995), combining aspects of computer graphics, human-computer interaction and data mining, the information visualization allows the presentation of data in a graphics form. In that way, the user can use their visual perception to better analyze and understand the information.

In general, the data do not have a direct, obvious and natural representation, which

contributes to the need of a good visual analysis technique. Such a technique in turn can use visual representations to present the data in a graphics form (LUZZARDI, 2003). In addition, according to Card, Mackinlay and Shneiderman (1999) the use of visual representations of interactive data with the support of the computer allows to increase the knowledge.

The information visualization is ideal for exploratory data analysis. Human eyes are naturally attracted to trends, patterns and exceptions that would be difficult to find using traditional approaches such as tables or text (FEW, 2009). The information visualization corresponds to the visual representation of abstract data to increase knowledge (SHIXIA et al., 2014). For example, representation of temporal statistics on the numbers of healthy children births, stock market trends, and information associated with the use of social networks. In general, interactive visual interfaces facilitate the process of expanding

knowledge and assist in the process of identifying patterns that are usually difficult to perceive.

2 Purpose of the Study

This study aims to relate and explore a large educational open dataset through information visualization techniques. For that, the database chosen was the microdata of the Brazilian national high school exam (ENEM), which has more than 7 million lines.

In order to investigate the relevance of information visualization tools, the Web application entitled VisDadosEnem has been developed so that the following research question can be answered: interactive information visualization tools are relevant in relating and exploring educational open data in a multidisciplinary context?

3 Methodology

In the elaboration of this study an applied and technological research was used. In order to obtain the data, the data portal of the Brazilian Federal Government was used, where open data is available from the most diverse public areas. And, for the Web application project, technologies were adopted that are more adequate to the process of treatment and visualization of educational data of the ENEM.

3.1 VisDadosEnem

Nowadays there is an increase in the data volume from the most different sources, as well as an increase in the open data made available by the Brazilian government (DADOS ABERTOS, 2016) and other governments (MÁCHOVÁ and LENICKA, 2017). Open data from several areas can be found, such as Education, Health, Technology, Defense, Security, Transportation, Traffic, Social Security, and Labor. With the data available in the administrative, technical and financial sectors of business and government organizations, there is a need for systems that allow to perform data analysis. Further, the analysis results make it possible to assist decision makers in extracting the information needed for decision-making process.

For the Web application project, it was carried out the data modeling and the

interface design. For data modeling, the conceptual model was first elaborated based on entity-relationship diagrams. In order to do that, only one table was needed to represent the candidate with their attributes. For that, it was included the enrollment number, the type of school attended, the state that resides, the marks in the knowledge areas, and the candidate's age based on 2014 ENEM microdata (DADOS ABERTOS, 2017). In addition, the logical model for the identification of the primary key, as well as the types and size of each attribute was elaborated. And, finally, for the description of the physical model, the table containing its attributes was created. Regarding the database, we chose MySQL[®] and for the Web application to be able to query the database, the PHP[®] language was adopted.

For the interface design, the conceptual design model, the prototype of the conceptual model and the detailed design of the user interface were developed. For that, it was taken into account functionalities associated to the data visualization methods to explore an adequate visual analysis. In this way, technologies for developing Web applications such as HTML[®], CSS[®] and Javascript[®] for the interface development were adopted. For the data visualization, the D3[®] (Data Driven Documents) library (D3, 2017) was used, which seeks to facilitate the understanding of the data by combining the interaction with the information visualization techniques. Finally, we used the NVD3[®] library (NVD3, 2017) that provides ready-made reusable graphics and graphic components for D3[®].

The information visualization techniques can be based on graphics, quantitative, hierarchical information, and number exploration. In addition, information visualization techniques can be classified according to a given set of data (WARD, GRINSTEIN, KEIN, 2015), such as: Multivariate data are those that do not usually have an explicit spatial attribute; Hierarchies and Trees are those that contain hierarchical characteristics and are considered a repository of data where there is a relationship with data subitems; Graphs and Networks are defined by their relationship characteristics based on graphs; And, Text and Documents are considered as being

literal, as a string, and defined by the set of objects as words, sentences, paragraphs, and documents.

The information visualization technique adopted for the visual representation of the educational data was the Multivariate data. The technique was based on a combination of elements where we can make use, for example of points, lines, and regions. We chose to develop a simple visualization with an easy understanding, and based on that, we decided to use bubbles (circles) for the visual representation. That type of visualization technique is known as bubble maps. The circle is a popularly known flat geometric figure, and, according to the variables to be displayed, the following properties have been added: area, color, and border color.

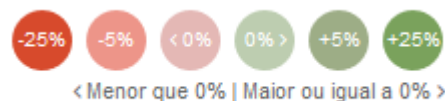
The circle area property is directly associated with the numeric value of the variable. The circle size is proportional to the value associated with the variable, as shown in Figure 1. In the case of VisDadosEnem, it indicates the number of candidates enrolled in the exam.

Figure 1: Circle area property. Prepared by the authors (2017).



The circle color property serves to classify the variable by intervals (Figure 2). The color property can assume two characteristic colors, where one color represents the lowest value and the other, the highest one. The interval between those values is represented by the variation of the intensity of each color. Higher values (positives or negatives) will have a more intense color tone. In the case of VisDadosEnem, it represents the comparison of the percentage of the average grade of a particular school type with a state or with Brazil.

Figure 2: Circle color property. Prepared by the authors (2017).



And, the border color property classifies the type that the variable can assume (Figure 3). Its color can vary according to the existing type, but it cannot be repeated in the same set, due to its unicity. In the case of VisDadosEnem, it represents the type of school (e.g., public, mostly public, private, and mostly private).

Figure 3: Border color property. Prepared by the authors (2017).



VisDadosEnem allows the visualization of the number of enrolments, the average grades by state or region, and the average age of the candidates. It also allows a comparative analysis of the averages obtained for each state or region by each type of school or area of knowledge through interactive visualization techniques. The Web application has the following main functionalities: Visualization by area of knowledge, by data between states, by types of schools, and by region¹.

The visualization by area of knowledge¹ allows visualizing the grades average obtained in all the states in the different types of school by area of knowledge. As shown in Appendix A, the interface allows the user to activate interactively select areas of knowledge for visualization through a check box menu.

The visualization of data between states¹ allows the comparison of grade averages obtained from one state compared to another state or to Brazil.

The visualization by type of schools¹¹ allows to analyze comparatively the performance of the types of school present in the states in function of the percentage of the grades average with respect to Brazil and the

¹ VisDadosEnem is available at: <http://labdata.sites.ufsc.br/visdadosenem>

average age of the candidates. The visualization occurs through bubble maps organized in a 2D chart (Appendix B) where the vertical axis represents the percentage of grades average for the different types of school and the horizontal axis represents the average age of the candidates.

And finally, the visualization by region¹ allows to analyze the performance of schools by region. The visualization also occurs through bubble maps organized in a 2D chart where the vertical axis represents the percentage of grades average for the regions of Brazil and the horizontal axis represents the average age of the candidates.

4 Results

Through the performed analysis with the information visualization techniques, it was possible to relate and explore the information available in VisDadosEnem.

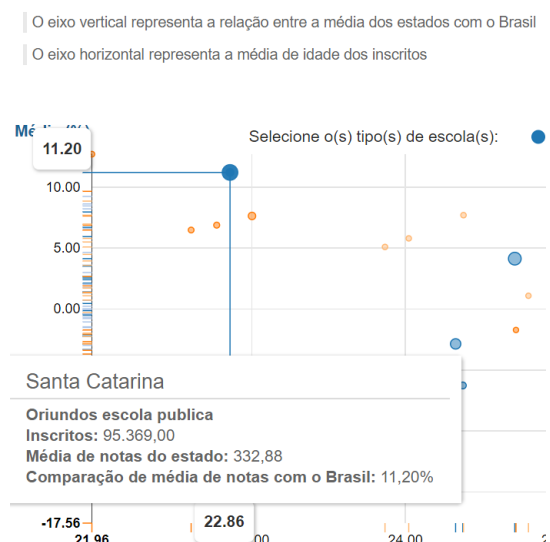
In the visualization feature by area of knowledge (Appendix A), it was possible to observe which states in Brazil, with students coming from public or mostly public schools, obtained an average grade higher than the national average compared to other states, as shown in Figure 4. That is, according to the bubble maps visualization the states that are above average in the results for public schools have circles, with the inner color in the green color variation and, respectively, present the border color in the blue color variation. It means that students are either in public schools (dark blue) or have attended mostly of the time in public schools (light blue).

Figure 4: A section of the visualization by area of knowledge. Prepared by the authors (2017).



In the visualization feature by type of schools (Appendix B), it was possible to observe, that the state of Santa Catarina had the highest grade average regarding students from private schools with 12.70% above the national average with the lowest average age of students enrolled corresponding to 21.6 years old. And, also, for students from public schools with 11.20% above national grade averages with the average age corresponding to 22.8 years old, as shown in Figure 5. On top of that, the lowest grade average regarding students from private schools was in the state of Amazonas with 17.56% below national grade averages with the average age corresponding to 26.8 years old. Also, for students from public schools, the state of Mato Grosso do Sul obtained the lowest grade average with 13.32% below the national average with the average age corresponding to 26.7 years old.

Figure 5: A section of the visualization by type of school. Prepared by the authors (2017).



4 Final Considerations

Although VisDadosEnem presents a set of basic functionalities for information visualization, the use of the visualization features made possible comparative analysis in the ENEM open data in order to relate and explore educational information. That type of exploratory analysis of information would be difficult to perform through tables or traditional graphics. In this way, with this

simple information visualization environment, one can see the relevance of a visualization tool for visual data exploration in a large volume of data.

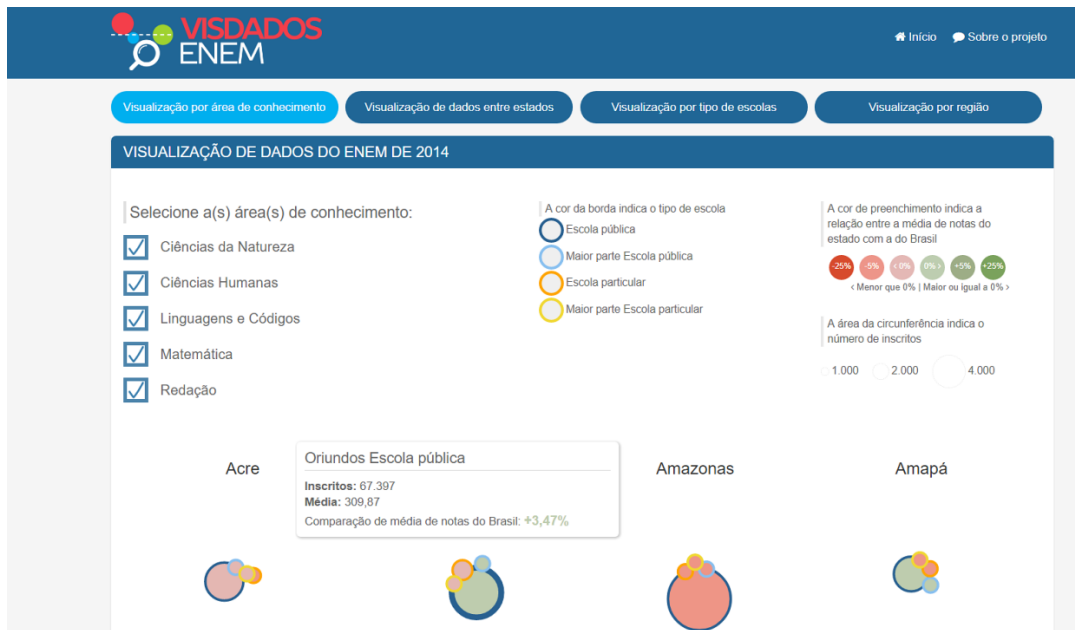
In addition, the study provided an online tool of public utility, which can be easily accessed and used for different types of visual analysis. It may be of interest to educational bodies, which can analyze results related to different types of schools, states, and regions. In this way, it can contribute to decision making process based on the information identified through the information visualization techniques.

In terms of future work, the educational open data used for the information visualization can be used in different scenarios. For example, it can be used to perform statistical studies, and to evaluate the quality of the public school in comparison to the private school in relation to the average grades obtained. Besides that, it is possible to identify which Brazilian states have a better or worse performance in the ENEM. Finally, as future work it would be interesting to extend the functionality of VisDadosEnem to allow the automatic insertion of updated ENEM data. In this way, as the data become available in the open data portal, it would be possible to perform temporal analysis as well as performance comparison between schools or states over time.

References

- CARD, K. S.; MACKINLAY, J. D.; SHNEIDERMAN, B. **Readings in Information Visualization, using vision to think**. Morgan Kaufmann, Cal. USA, 1999.
- DADOS ABERTOS. **Portal Brasileiro de Dados Abertos**. Available at: <<http://dados.gov.br/>>. Accessed: 22 Jul. 2017.
- D3. **Data-Driven Documents**. Available at: <<http://d3js.org/>>. Accessed: 22 jul. 2017.
- FEW, S. **Now you see it: simple visualization techniques for quantitative analysis, Analytics**. Press, 2009.
- FREITAS, C. M. D. S; WAGNER, F. R. Suporte às tarefas da análise exploratória visual. **Revista de Informática Teórica e Aplicada**, vol. 2, n.1, p. 5-36, jan. 1995.
- LUZZARDI, P. R. G. **Critérios de Avaliação de Técnicas de Visualização de Informações Hierárquicas**. Tese de doutorado (Programa de Pós-Graduação em Computação), UFRGS, Porto Alegre, 2003.
- MÁCHOVÁ, R.; LNENICKA, M. Evaluating the Quality of Open Data Portals on the National Level. **Journal of Theoretical and Applied Electronic Commerce Research**, vol. 12, n. 1, p. 21-41, jan. 2017.
- NVD3. **Re-usable Charts for D3**. Available at: <<http://nvd3.org/>>. Accessed: 22 jul. 2017.
- SHIXIA, L.; WEIWEI, C.; YINGCAI, W.; MENGCHEN, L. A survey on information visualization: recent advances and challenges. **The Visual Computer**, vol. 30, n. 12, p. 1373-1393, 2014.
- WARD, M.; GRINSTEIN, G.; KEIN, D. **Interactive data visualization: foundations, techniques, and applications**. CRC Press, 2015.

Appendix A - Visualization by area of knowledge. Exploratory analysis of the information through visualization by area of knowledge. The check boxes for each knowledge area can be enabled or disabled. Prepared by the authors (2017).



Appendix B – Visualization by type of schools. Exploratory analysis of the information through visualization by type of school. Each state has a representation for each of the 4 types of schools (i.e., blue circle - public school, blue light - mostly public school, orange - private school, and light orange - mostly private school). The circles for each type of school can be enabled or disabled. Prepared by the authors (2017).

